Our Ref.: P18245

UNITED STATES PATENT APPLICATION

FOR

A METHOD FOR MAKING A SEMICONDUCTOR DEVICE THAT INCLUDES A METAL GATE ELECTRODE

Inventors:

Mark L. Doczy
Justin K. Brask
Jack Kavalieros
Uday Shah
Chris E. Barns
Robert S. Chau

Attorney's Docket No.: P18245

"Express Mail" mailing label number: EV 409362025
Date of Deposit: December 29, 2003
I hereby state that I am causing this paper or fee to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated above and that this paper or fee has been addressed to the Commissioner for Patents, PO Box 1450, Alexandria, Virginia 22313-1450

TERESA Edwards
(Typed or printed name of person mailing paper or fee)

Teresa Edwards
(Signature of person mailing paper or fee)

December 29, 2003
(Date signed)

A METHOD FOR MAKING A SEMICONDUCTOR DEVICE THAT INCLUDES A METAL GATE ELECTRODE

FIELD OF THE INVENTION

[0001] The present invention relates to methods for making semiconductor devices, in particular, semiconductor devices with metal gate electrodes.

BACKGROUND OF THE INVENTION

[0002] CMOS devices with very thin gate dielectrics made from silicon dioxide may experience unacceptable gate leakage currents. Forming the gate dielectric from certain high-k dielectric materials, instead of silicon dioxide, can reduce gate leakage. Because, however, such a dielectric may not be compatible with polysilicon, it may be desirable to replace polysilicon based gate electrodes with metal gate electrodes in devices that include high-k gate dielectrics.

[0003] To form metal NMOS and PMOS gate electrodes that have appropriate workfunctions, it may be necessary to form them from different materials – one that ensures an acceptable workfunction for the NMOS gate electrode, and another that ensures an acceptable workfunction for the PMOS gate electrode. A replacement gate process may be used to form metal NMOS and PMOS gate electrodes from different metals. In that process, a first polysilicon layer, bracketed by a pair of spacers, is removed selectively to a second polysilicon layer to create a trench between the spacers. The trench is filled with a first metal. The second polysilicon layer is then removed, and replaced with a second metal that differs from the first metal.

[0004] When using such a replacement gate process to form metal NMOS and PMOS gate electrodes, it may be necessary to form a hard mask on the polysilicon layers to minimize silicide formation, when the transistors' source and drain regions are covered with a silicide. Although such a hard mask may protect the upper surface of the polysilicon

layers, the upper corners of those layers may be exposed, when the spacers are formed. Silicide may form at those exposed corners, when the source and drain regions are silicided, which may adversely impact the subsequent polysilicon removal steps.

[0005] Accordingly, there is a need for an improved method for making a semiconductor device that includes metal gate electrodes. There is a need for a replacement gate process that replaces polysilicon layers with metal layers, which is not adversely affected by silicide formation on the polysilicon layers. The present invention provides such a method.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] Figures 1a-1k represent cross-sections of structures that may be formed when carrying out an embodiment of the method of the present invention.

[0007] Figures 2a-2c represent cross-sections of structures that may be formed when carrying out an alternative embodiment of the method of the present invention.

[0008] Features shown in these figures are not intended to be drawn to scale.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0009] A method for making a semiconductor device is described. That method comprises forming a patterned sacrificial gate electrode layer that is covered by a hard mask and an etch stop layer. After forming first and second spacers on opposite sides of the patterned sacrificial gate electrode layer, that layer is removed to generate a trench that is positioned between the first and second spacers. At least part of that trench is then filled with a metal layer. In the following description, a number of details are set forth to provide a thorough understanding of the present invention. It will be apparent to those skilled in the art, however, that the invention may be practiced in many ways other than those expressly described here. The invention is thus not limited by the specific details disclosed below.

[0010] Figures 1a-1k illustrate structures that may be formed, when carrying out an embodiment of the method of the present invention. Figure 1a represents an intermediate structure that may be formed when making a complementary metal oxide semiconductor ("CMOS"). That structure includes first part 101 and second part 102 of substrate 100. Isolation region 103 separates first part 101 from second part 102. First sacrificial gate electrode layer 104 is formed on first gate dielectric layer 105, and second sacrificial gate electrode layer 106 is formed on second gate dielectric layer 107. Hard masks 130, 131 are formed on sacrificial gate electrode layers 104, 106, and etch stop layers 132, 133 are formed on hard masks 130, 131.

[0011] Substrate 100 may comprise a bulk silicon or silicon-on-insulator substructure. Alternatively, substrate 100 may comprise other materials -- which may or may not be combined with silicon -- such as: germanium, indium antimonide, lead telluride, indium arsenide, indium phosphide, gallium arsenide, or gallium antimonide. Although a few examples of materials from which substrate 100 may be formed are described here, any material that may serve as a foundation upon which a semiconductor device may be built falls within the spirit and scope of the present invention.

[0012] Isolation region 103 may comprise silicon dioxide, or other materials that may separate the transistor's active regions. First gate dielectric layer 105 and second gate dielectric layer 107 may each comprise silicon dioxide, or other materials that may insulate the substrate from the gate electrodes. Dielectric layers 105, 107 preferably comprise a high quality, dense thermally grown silicon dioxide layer that is less than about 20 angstroms thick, and more preferably that is about 10 angstroms thick. Sacrificial gate electrode layers 104, 106 may comprise polysilicon and are preferably between about 100

and about 2,000 angstroms thick, and more preferably between about 500 and about 1,600 angstroms thick.

[0013] Hard masks 130, 131 may comprise silicon nitride, and preferably are between about 100 and about 500 angstroms thick -- and more preferably between about 200 and about 350 angstroms thick. Etch stop layers 132, 133 may comprise a material that will be removed at a substantially slower rate than silicon nitride will be removed when an appropriate etch process is applied. Etch stop layers 132, 133 may, for example, be made from silicon, an oxide (e.g., silicon dioxide or a metal oxide such as hafnium dioxide), a carbide (e.g., silicon carbide), or a carbon doped silicon nitride. Etch stop layers 132, 133 preferably are between about 200 and about 1,200 angstroms thick, and more preferably are between about 400 and about 600 angstroms thick.

[0014] When sacrificial gate electrode layers 104, 106 comprise polysilicon, and hard mask layers 130, 131 comprise silicon nitride, the figure 1a structure may be made in the following way. A dielectric layer, which may comprise silicon dioxide, is formed on substrate 100 (e.g., via a conventional thermal growth process), followed by forming a polysilicon layer on the dielectric layer (e.g., via a conventional deposition process). Using conventional deposition techniques, a silicon nitride layer is formed on the polysilicon layer, and an etch stop layer is formed on the silicon nitride layer. The etch stop, silicon nitride, polysilicon, and dielectric layers are then patterned – using conventional lithography and etch processes – to form patterned etch stop layers 132, 133, patterned silicon nitride layers 130, 131, patterned polysilicon layers 104, 106, and patterned dielectric layers 105, 107.

[0015] After forming the figure 1a structure, spacers are formed on opposite sides of sacrificial gate electrode layers 104, 106. When those spacers comprise silicon nitride,

they may be formed in the following way. First, a silicon nitride layer of substantially uniform thickness -- preferably less than about 1000 angstroms thick -- is deposited over the entire structure, producing the structure shown in figure 1b. Conventional deposition processes may be used to generate that structure.

[0016] In a preferred embodiment, silicon nitride layer 134 is deposited directly on substrate 100, patterned etch stop layers 132, 133, and opposite sides of sacrificial gate electrode layers 104, 106 -- without first forming a buffer oxide layer on substrate 100 and layers 104, 106. In alternative embodiments, however, such a buffer oxide layer may be formed prior to forming layer 134. Similarly, although not shown in figure 1b, a second oxide may be formed on layer 134 prior to etching that layer. If used, such an oxide may enable the subsequent silicon nitride etch step to generate an L-shaped spacer.

[0017] Silicon nitride layer 134 may be etched using a conventional process for anisotropically etching silicon nitride to create the figure 1c structure. Etch stop layers 132, 133 prevent such an anisotropic etch step from removing hard masks 130, 131, when silicon nitride layer 134 is etched – even when hard masks 130, 131 comprise silicon nitride. As a result of that etch step, sacrificial gate electrode layer 104 is bracketed by a pair of sidewall spacers 108, 109, and sacrificial gate electrode layer 106 is bracketed by a pair of sidewall spacers 110, 111.

[0018] As is typically done, it may be desirable to perform multiple masking and ion implantation steps to create lightly implanted regions near layers 104, 106 (that will ultimately serve as tip regions for the device's source and drain regions), prior to forming spacers 108, 109, 110, 111 on sacrificial gate electrode layers 104, 106. Also as is typically done, the source and drain regions may be formed, after forming spacers 108, 109, 110, 111, by implanting ions into parts 101 and 102 of substrate 100, followed by

applying an appropriate anneal step. Part of those source and drain regions may then be converted to a silicide using well known process steps. Etch stop layers 132, 133 will prevent such a process sequence from converting a meaningful part, if any, of sacrificial gate electrode layers 104, 106 to a silicide.

[0019] When sacrificial gate electrode layers 104, 106 comprise polysilicon, an ion implantation and anneal sequence used to form n-type source and drain regions within part 101 of substrate 100 may dope polysilicon layer 104 n-type at the same time. Similarly, an ion implantation and anneal sequence used to form p-type source and drain regions within part 102 of substrate 100 may dope polysilicon layer 106 p-type. When doping polysilicon layer 106 with boron, that layer should include that element at a sufficient concentration to ensure that a subsequent wet etch process, for removing n-type polysilicon layer 104, will not remove a significant amount of p-type polysilicon layer 106.

[0020] After forming spacers 108, 109, 110, 111, dielectric layer 112 may be deposited over the device, generating the figure 1d structure. Dielectric layer 112 may comprise silicon dioxide, or a low-k material. Dielectric layer 112 may be doped with phosphorus, boron, or other elements, and may be formed using a high density plasma deposition process. By this stage of the process, source and drain regions 135, 136, 137, 138, which are capped by silicided regions 139, 140, 141, 142, have already been formed. Conventional process steps, materials, and equipment may be used to generate the structures represented by figures 1a-1d, as will be apparent to those skilled in the art. Those structures may include other features -- not shown, so as not to obscure the method of the present invention -- that may be formed using conventional process steps.

[0021] Dielectric layer 112 is removed from patterned etch stop layers 132, 133, which are, in turn, removed from hard masks 130, 131, which are, in turn, removed from patterned

sacrificial gate electrode layers 104, 106, producing the figure 1e structure. A conventional chemical mechanical polishing ("CMP") operation may be applied to remove that part of dielectric layer 112, patterned etch stop layers 132, 133, and hard masks 130, 131. Etch stop layers 132, 133 and hard masks 130, 131 must be removed to expose patterned sacrificial gate electrode layers 104, 106. Etch stop layers 132, 133 and hard masks 130, 131 may be polished from the surface of layers 104, 106, when dielectric layer 112 is polished – as they will have served their purpose by that stage in the process.

[0022] After forming the figure 1e structure, sacrificial gate electrode layer 104 is removed to generate trench 113 that is positioned between sidewall spacers 108, 109 -- producing the structure shown in figure 1f. In a preferred embodiment, a wet etch process that is selective for layer 104 over sacrificial gate electrode layer 106 is applied to remove layer 104 without removing significant portions of layer 106.

[0023] When sacrificial gate electrode layer 104 is either undoped or doped n-type, and sacrificial gate electrode layer 106 is doped p-type (e.g., with boron), such a wet etch process may comprise exposing sacrificial gate electrode layer 104 to an aqueous solution that comprises a source of hydroxide for a sufficient time at a sufficient temperature to remove substantially all of layer 104. That source of hydroxide may comprise between about 2 and about 30 percent ammonium hydroxide or a tetraalkyl ammonium hydroxide, e.g., tetramethyl ammonium hydroxide ("TMAH"), by volume in deionized water.

[0024] Sacrificial gate electrode layer 104 may be selectively removed by exposing it to a solution, which is maintained at a temperature between about 15°C and about 90°C (and preferably below about 40°C), that comprises between about 2 and about 30 percent ammonium hydroxide by volume in deionized water. During that exposure step, which preferably lasts at least one minute, it may be desirable to apply sonic energy at a

frequency of between about 10 KHz and about 2,000 KHz, while dissipating at between about 1 and about 10 watts/cm^2.

[0025] In a particularly preferred embodiment, sacrificial gate electrode layer 104, with a thickness of about 1,350 angstroms, may be selectively removed by exposing it at about 25°C for about 30 minutes to a solution that comprises about 15 percent ammonium hydroxide by volume in deionized water, while applying sonic energy at about 1,000 KHz -- dissipating at about 5 watts/cm^2. Such an etch process should remove substantially all of an n-type polysilicon layer without removing a meaningful amount of a p-type polysilicon layer.

[0026] As an alternative, sacrificial gate electrode layer 104 may be selectively removed by exposing it for at least one minute to a solution, which is maintained at a temperature between about 60°C and about 90°C, that comprises between about 20 and about 30 percent TMAH by volume in deionized water, while applying sonic energy. Removing sacrificial gate electrode layer 104, with a thickness of about 1,350 angstroms, by exposing it at about 80°C for about 2 minutes to a solution that comprises about 25 percent TMAH by volume in deionized water, while applying sonic energy at about 1,000 KHz -- dissipating at about 5 watts/cm^2 -- may remove substantially all of layer 104 without removing a significant amount of layer 106.

[0027] In this embodiment, after removing sacrificial gate electrode layer 104, first gate dielectric layer 105, which may comprise silicon dioxide, is retained -- followed by forming n-type metal layer 115 on layer 105 to fill trench 113 and to generate the figure 1g structure. N-type metal layer 115 may comprise any n-type conductive material from which a metal NMOS gate electrode may be derived. N-type metal layer 115 preferably has

thermal stability characteristics that render it suitable for making a metal NMOS gate electrode for a semiconductor device.

[0028] Materials that may be used to form n-type metal layer 115 include: hafnium, zirconium, titanium, tantalum, aluminum, and their alloys, e.g., metal carbides that include these elements, i.e., hafnium carbide, zirconium carbide, titanium carbide, tantalum carbide, and aluminum carbide. N-type metal layer 115 may be formed on first gate dielectric layer 105 using well known PVD or CVD processes, e.g., conventional sputter or atomic layer CVD processes. As shown in figure 1h, n-type metal layer 115 is removed except where it fills trench 113. Layer 115 may be removed from other portions of the device via a wet or dry etch process, or an appropriate CMP operation. Dielectric 112 may serve as an etch or polish stop, when layer 115 is removed from its surface.

[0029] N-type metal layer 115 preferably serves as a metal NMOS gate electrode that has a workfunction that is between about 3.9 eV and about 4.2 eV, and that is between about 100 angstroms and about 2,000 angstroms thick, and more preferably is between about 500 angstroms and about 1,600 angstroms thick. Although figures 1g and 1h represent structures in which n-type metal layer 115 fills all of trench 113, in alternative embodiments, n-type metal layer 115 may fill only part of trench 113, with the remainder of the trench being filled with a material that may be easily polished, e.g., tungsten or aluminum. In such an alternative embodiment, n-type metal layer 115, which serves as the workfunction metal, may be between about 50 and about 1,000 angstroms thick – and more preferably at least about 100 angstroms thick.

[0030] In embodiments in which trench 113 includes both a workfunction metal and a trench fill metal, the resulting metal NMOS gate electrode may be considered to comprise the combination of both the workfunction metal and the trench fill metal. If a trench fill

metal is deposited on a workfunction metal, the trench fill metal may cover the entire device when deposited, forming a structure like the figure 1g structure. That trench fill metal must then be polished back so that it fills only the trench, generating a structure like the figure 1h structure.

[0031] In the illustrated embodiment, after forming n-type metal layer 115 within trench 113, sacrificial gate electrode layer 106 is removed to generate trench 150 that is positioned between sidewall spacers 110, 111 -- producing the structure shown in figure 1i. In a preferred embodiment, layer 106 is exposed to a solution that comprises between about 20 and about 30 percent TMAH by volume in deionized water for a sufficient time at a sufficient temperature (e.g., between about 60°C and about 90°C), while applying sonic energy, to remove all of layer 106 without removing significant portions of n-type metal layer 115.

[0032] Alternatively, a dry etch process may be applied to selectively remove layer 106. When sacrificial gate electrode layer 106 is doped p-type (e.g., with boron), such a dry etch process may comprise exposing sacrificial gate electrode layer 106 to a plasma derived from sulfur hexafluoride ("SF_6"), hydrogen bromide ("HBr"), hydrogen iodide ("HI"), chlorine, argon, and/or helium. Such a selective dry etch process may take place in a parallel plate reactor or in an electron cyclotron resonance etcher.

[0033] In this embodiment, after removing sacrificial gate electrode layer 106, second gate dielectric layer 107, which may comprise silicon dioxide, is retained -- followed by forming p-type metal layer 116 on layer 107 to fill trench 150 and to generate the figure 1j structure. P-type metal layer 116 may comprise any p-type conductive material from which a metal PMOS gate electrode may be derived. P-type metal layer 116 preferably has thermal

stability characteristics that render it suitable for making a metal PMOS gate electrode for a semiconductor device.

[0034] Materials that may be used to form p-type metal layer 116 include: ruthenium, palladium, platinum, cobalt, nickel, and conductive metal oxides, e.g., ruthenium oxide. P-type metal layer 116 may be formed on second gate dielectric layer 107 using well known PVD or CVD processes, e.g., conventional sputter or atomic layer CVD processes. As shown in figure 1k, p-type metal layer 116 is removed except where it fills trench 150. Layer 116 may be removed from other portions of the device via a wet or dry etch process, or an appropriate CMP operation, with dielectric 112 serving as an etch or polish stop.

[0035] P-type metal layer 116 may serve as a metal PMOS gate electrode with a workfunction that is between about 4.9 eV and about 5.2 eV, and that is between about 100 angstroms and about 2,000 angstroms thick, and more preferably is between about 500 angstroms and about 1,600 angstroms thick.

[0036] Although figures 1j and 1k represent structures in which p-type metal layer 116 fills all of trench 150, in alternative embodiments, p-type metal layer 116 may fill only part of trench 150. As with the metal NMOS gate electrode, the remainder of the trench may be filled with a material that may be easily polished, e.g., tungsten or aluminum. In such an alternative embodiment, p-type metal layer 116, which serves as the workfunction metal, may be between about 50 and about 1,000 angstroms thick. Like the metal NMOS gate electrode, in embodiments in which trench 150 includes a workfunction metal and a trench fill metal, the resulting metal PMOS gate electrode may be considered to comprise the combination of both the workfunction metal and the trench fill metal.

[0037] Although a few examples of materials that may be used to form metal layers 115 and 116 are described here, those metal layers may be made from many other materials,

as will be apparent to those skilled in the art. After removing metal layer 116, except where it fills trench 150, a capping dielectric layer may be deposited onto dielectric layer 112, metal NMOS gate electrode 115, and metal PMOS gate electrode 116, using any conventional deposition process. Process steps for completing the device that follow the deposition of such a capping dielectric layer, e.g., forming the device's contacts, metal interconnect, and passivation layer, are well known to those skilled in the art and will not be described here.

[0038] Figures 1a-1k illustrate an embodiment in which first and second gate dielectric layers 105, 107 are retained after layers 104, 106 are removed. Figures 2a-2c illustrate an alternative embodiment, in which first gate dielectric layer 105 is removed after removing sacrificial gate electrode layer 104, and second gate dielectric layer 107 is removed after removing sacrificial gate electrode layer 106. Figure 2a represents the device after layer 104 is removed. That structure may be identical to the figure 1f structure. In this alternative embodiment, first gate dielectric layer 105 is removed after sacrificial gate electrode layer 104 is removed.

[0039] When first gate dielectric layer 105 comprises silicon dioxide, it may be removed using an etch process that is selective for silicon dioxide to generate the figure 2b structure. Such etch processes include: exposing layer 105 to a solution that includes about 1 percent HF in deionized water, or applying a dry etch process that employs a fluorocarbon based plasma. Layer 105 should be exposed for a limited time, as the etch process for removing layer 105 may also remove part of dielectric layer 112. With that in mind, if a 1 percent HF based solution is used to remove layer 105, the device preferably should be exposed to that solution for less than about 60 seconds, and more preferably for about 30 seconds or less.

[0040] When first gate dielectric layer 105 is removed, it must be replaced prior to forming an n-type metal layer within trench 113. Preferably, high-k gate dielectric layer 114 is formed on substrate 100 at the bottom of trench 113, after first gate dielectric layer 105 has been removed -- generating the structure illustrated by figure 2c. Some of the materials that may be used to make high-k gate dielectric 114 include: hafnium oxide, hafnium silicon oxide, lanthanum oxide, zirconium oxide, zirconium silicon oxide, tantalum oxide, barium strontium titanium oxide, barium titanium oxide, strontium titanium oxide, yttrium oxide, aluminum oxide, lead scandium tantalum oxide, and lead zinc niobate. Particularly preferred are hafnium oxide, zirconium oxide, and aluminum oxide. Although a few examples of materials that may be used to form high-k gate dielectric layer 114 are described here, that layer may be made from other materials.

[0041] High-k gate dielectric layer 114 may be formed on substrate 100 using a conventional deposition method, e.g., a conventional chemical vapor deposition ("CVD"), low pressure CVD, or physical vapor deposition ("PVD") process. Preferably, a conventional atomic layer CVD process is used. In such a process, a metal oxide precursor (e.g., a metal chloride) and steam may be fed at selected flow rates into a CVD reactor, which is then operated at a selected temperature and pressure to generate an atomically smooth interface between substrate 100 and high-k gate dielectric layer 114. The CVD reactor should be operated long enough to form a layer with the desired thickness. In most applications, high-k gate dielectric layer 114 should be less than about 60 angstroms thick, and more preferably between about 5 angstroms and about 40 angstroms thick.

[0042] Although not shown in figure 2c, if an atomic layer CVD process is used to form high-k gate dielectric layer 114, that layer may form on the sides of trench 113 in addition

to forming on the bottom of that trench. If high-k gate dielectric layer 114 comprises an oxide, it may manifest oxygen vacancies at random surface sites and unacceptable impurity levels, depending upon the process used to make it. It may be desirable to remove certain impurities from layer 114, and to oxidize it to generate a layer with a nearly idealized metal:oxygen stoichiometry, after layer 114 is deposited.

[0043] Likewise, second gate dielectric layer 107 may be replaced with a high-k dielectric layer after sacrificial gate electrode layer 106 is removed and before p-type metal layer 116 fills trench 150. The method described above enables production of CMOS devices that include metal NMOS and PMOS gate electrodes without causing significant amounts of silicide to form on the polysilicon layers, which must be removed prior to forming the metal gate electrodes. Although the embodiments described above provide examples of processes for forming such devices, the present invention is not limited to these particular embodiments.

[0044] Although the foregoing description has specified certain steps and materials that may be used in the present invention, those skilled in the art will appreciate that many modifications and substitutions may be made. Accordingly, it is intended that all such modifications, alterations, substitutions and additions be considered to fall within the spirit and scope of the invention as defined by the appended claims.